УДК 621.391.23

Ю. Г. Горшков

ИССЛЕДОВАТЕЛЬСКИЙ КОМПЛЕКС ЧАСТОТНО-ВРЕМЕННОГО АНАЛИЗА РЕЧЕВОГО СИГНАЛА С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТ-ТЕХНОЛОГИИ

Рассмотрены недостатки распространенных аппаратно-программных средств анализа речи, используемых при экспертизе фонограмм. Представлена структура исследовательского комплекса частотно-временного анализа речевого сигнала, обеспечивающего повышенную точность обработки речевой информации с использованием вейвлет-технологии. Приведены экспериментальные данные построения вейвлет-сонограмм гласных и согласных звуков.

E-mail: ygorshkov@rambler.ru

Ключевые слова: *речевой сигнал, частотно-временной анализ, вейвлет-преобразование.*

В последние годы анализ аудиозаписей звуковой или речевой информации находит все большее применение как в государственных экспертных учреждениях [1], так и в частных охранных структурах. Широкое применение малогабаритных средств регистрации, позволяющих осуществлять цифровую запись речевой информации в сложной акустической обстановке, в том числе в условиях противодействия звукозаписи, определяет задачи высокоточного анализа и очистки речевого сигнала от шумов и помех как наиболее актуальные [2].

Надежность систем распознавания слитной устной речи и идентификации диктора по голосу также зависит от точности методов и алгоритмов выделения информационных параметров речи при статистической обработке акустических сигналов. Достаточно сложной задачей является поиск объективных характеристик физиологических закономерностей образования звуков в различных языках, определяемых общими принципами формирования речевого звучания, для которых первичным является наличие около 50 звуков речи, разделяемых на гласные и согласные.

Большинство современных методов анализа звуков речи основаны на спектральной модели стационарного сигнала [3]. Недостатком этой модели является отсутствие вероятностных характеристик для основных шумовых составляющих в произносимых согласных, и это при

том, что в большинстве языков основная речевая информация передается согласными звуками. Так, в русском языке из 43 основных звуков — 6 гласных и 37 согласных.

В таблице представлены данные о соотношении гласных и согласных звуков для иностранных языков.

Таблица

Соотношение гласных и согласных звуков в разных языках

Звуки	Язык				
	Английский	Арабский	Немецкий	Французский	Японский
Гласные	20	3	18	16	5
Согласные	24	28	24	17	26

Традиционно разрабатываемые алгоритмы распознавания речи и идентификации личности по голосу основываются на оценке значений основного тона или формант гласных звуков. Количественные параметры формант используются для поиска отличий между звуками. Обычно наиболее информативными считаются первые две форманты, а для поиска личностных признаков анализируют третья–пятая форманты [4].

Аппаратно-программные комплексы исследования фонограмм. Аппаратно-программные комплексы исследования фонограмм используются экспертами-криминалистами при решении задач анализа речевых сигналов и идентификации дикторов. Широкое применение находят комплексы, разработанные российскими компаниями: “Диалект”, “ИкарЛаб”, Justiphone, OTExpert. Идентификация личности по голосу основана на оценке значений основного тона или формант гласных звуков, вычисленных по алгоритму быстрого преобразования Фурье (БПФ).

На этапе предварительного анализа фонограмм используются звуковые редакторы: Adobe Audition, AWave, Cool Edit, Sound Forge, Speech Analyzer, Steinberg WaveLab, Wave Flow, WaveLab. Цифровая обработка сигналов с использованием перечисленных звуковых редакторов также основана на алгоритмах БПФ.

На рис. 1 приведена сонограмма или частотно-временное представление триады звуков “н эс”; значения периода основного тона и формант получены с использованием комплекса “Диалект”.

Вычисление акустических признаков при микроанализе звуков диктора по существующей методике [5] проводится на наиболее информативных, с точки зрения проявления индивидуальности, гласных звуках [а], [о], [е], [и]. Основными параметрами, характеризующими индивидуальность голоса диктора для сопоставимых по контексту звуков, считаются:

— значения частоты основного тона (F0) на гласных звуках;

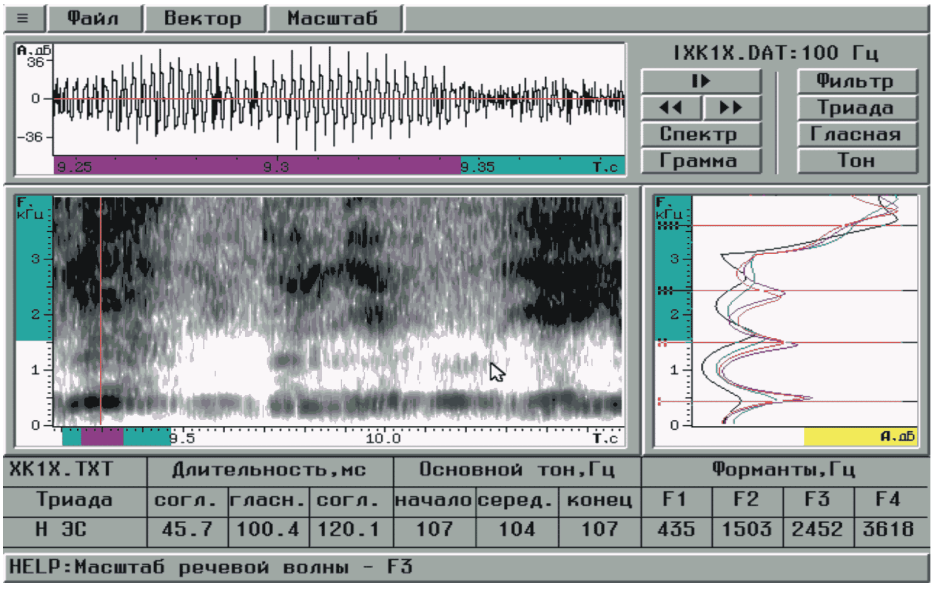

Рис. 1. Фурье-сонограмма, значения основного тона и формант триады звуков “н эс”

— значения четырех формантных частот (F1, F2, F3, F4) гласных звуков;

— длительность гласного звука (Тг);

— длительность согласных, окружающих гласный звук (Тс).

Частота F0 связана с индивидуальными физиологическими характеристиками голосовых связок говорящего.

Для определения индивидуальной для каждого диктора вариации частоты основного тона на гласных значение F0 вычисляется в начале соответствующей гласной, в средней части и в конце.

Формантные частоты F1–F4 — это первые четыре (по порядку) резонансные частоты спектров гласных звуков. На этих частотах концентрируется подавляющая часть энергетического спектра гласных. Частоты формант отражают индивидуальные физиологические параметры речеобразующих органов говорящего.

Таким образом, комплексом “Диалект”, а также остальными известными аппаратно-программными средствами криминалистического исследования фонограмм определяются акустические признаки только гласных звуков, в согласных звуках анализируется только их длительность. Эти ограничения вызваны недостатками преобразования Фурье при обработке нестационарных сигналов:

— исходный сигнал заменяется на периодический с периодом, равным длительности анализируемого участка;

— преобразование Фурье не обеспечивает необходимую точность при изменении параметров процесса во времени (нестационарности),

поскольку дает усредненные коэффициенты для всего исследуемого сигнала.

Для выполнения анализа нестационарного процесса необходимо использовать базисные функции, имеющие способность выявлять в анализируемом сигнале как частотные, так и его временные характеристики. Другими словами, сами функции должны обладать свойствами частотно-временной локализации.

В большей степени перечисленным требованиям отвечают *вейвлеты* и, соответственно, математическим методам анализа — *вейвлет-преобразование.*

Обработка речевых сигналов с использованием вейвлет-преобразования. В последние годы вейвлеты находят широкое применение при фильтрации, предварительной обработке и синтезе различных сигналов, решении задач сжатия и обработки изображений. Вейвлеты были предложены математиками и по существу являются новыми математическими понятиями и объектами. Особенно важна принципиальная возможность вейвлетов представлять нестационарные сигналы [6]. Все большее число специалистов по цифровой обработке сигналов убеждаются в том, что преобразования Фурье в классическом виде не обеспечивают необходимую точность представления нестационарных сигналов, к которым, в частности, относятся речевые сигналы. Значительный интерес представляют утверждения о том, что вейвлет-спектрограммы намного более информативны, чем обычные фурье-спектрограммы, и в отличие от последних позволяют выявлять тончайшие локальные особенности акустических сигналов. На рис. 2 приведена структура представления сигнала при использовании вейвлет-преобразования и преобразования Фурье. Очевидно, что вейвлет-преобразование отличается более сложной и гибкой структурой обработки.

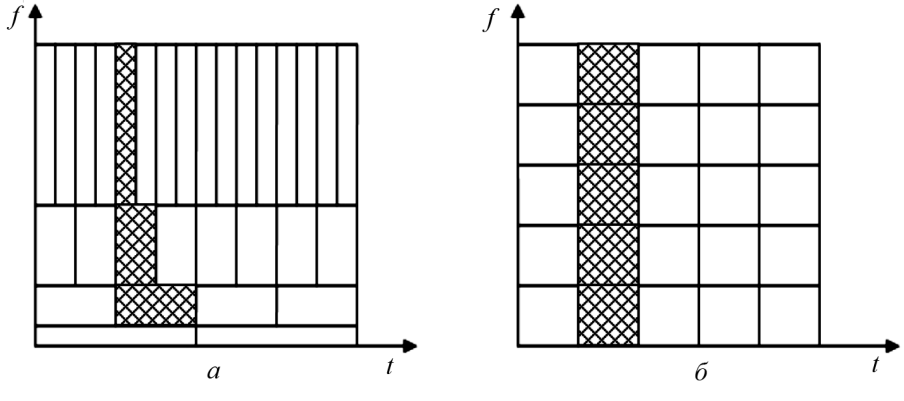

Рис. 2. Структура представления сигнала при вейвлет-преобразовании (*а*) и при преобразовании Фурье (*б*)

Вейвлет-преобразование сигналов является обобщением спектрального анализа. Применяемые для этой цели базисы были названы вейвлетами, т.е. функциями двух аргументов — масштаба и сдвига. В отличие от традиционного преобразования Фурье, вейвлет-преобразование обеспечивает двумерное представление исследуемого сигнала в частотной области в плоскости частота–положение. Аналогом частоты при этом является масштаб аргумента базисной функции (чаще всего — времени), а положение характеризуется ее сдвигом. Это позволяет разделять крупные и мелкие особенности сигналов, одновременно локализуя их на временно́й шкале. Иными словами, вейвлет-анализ можно охарактеризовать как спектральный анализ локальных возмущений [7].

Непрерывный вейвлет-анализ. Результатом непрерывного вейвлет-анализа некоторого сигнала, заданного функцией $f(t)$, будет функция $Wf(a,b)$, которая зависит уже от двух переменных — от координаты b и масштаба a:

$$Wf(ab) = \frac{1}{\sqrt{a}} \int\limits_{-\infty}^{\infty} \psi\Big(\frac{t-b}{a}\Big) f(t)dt. \qquad (1)$$

Распределение значений коэффициентов $Wf(a,b)$ в пространстве (a,b) дает информацию о вкладе компонент во времени и называется спектром коэффициентов вейвлет-преобразования, (частотно-) масштабно-временным спектром или вейвлет-спектром.

Способы визуализации этой информации могут быть различными. Наибольшее распространение получило представление спектра $Wf(a,b)$ в виде проекции на плоскость (a,b) с изолиниями, позволяющими проследить изменения интенсивности амплитуд вейвлет-преобразования на разных масштабах и во времени, а также картины линий локальных экстремумов этих поверхностей, четко выявляющие структуру анализируемого процесса.

Многомасштабный вейвлет-анализ. Многомасштабный (кратномасштабный) вейвлет-анализ является развитием дискретного вейвлет-анализа (сущность которого состоит в представлении сигнала последовательностью образов с разной степенью детализации, что позволяет выявлять локальные особенности сигнала и классифицировать их по интенсивности) и основывается на разложении сигнала по функциям, образующим ортонормированный базис [8]. Любую функцию можно разложить на некотором заданном уровне разрешения (масштабе) j_n в ряд вида

$$f(t) = \sum_{k=0}^{2M-1} s_{j_n,k}\varphi_{j_n,k} + \sum_{j\geq j_n}^{j_{\max}} \sum_{k=0}^{2M-1} d_{j_n,k}\psi_{j,k}. \qquad (2)$$

Здесь $\varphi_{j_n,k}$ и $\psi_{j,k}$ — масштабированные и смещенные версии скейлинг-функции (масштабной функции) φ и материнского вейвлета ψ; $s_{j,k}$ — коэффициенты аппроксимации; $d_{j,k}$ — детализирующие коэффициенты.

Масштабирование и смещение функций φ и ψ находится следующим образом:

$$\varphi_{j,k} = 2^{jj/2}\varphi(2^j t - k); \tag{3}$$

$$\psi_{j,k} = 2^{j/2}\psi(2^j t - k). \tag{4}$$

В свою очередь сами функции φ и ψ определяются как

$$\varphi(t) = \sqrt{2}\sum_{k=0}^{2M-1} h_k \varphi(2t - k); \tag{5}$$

$$\psi(t) = \sqrt{2}\sum_{k=0}^{2M-1} g_k \varphi(2t - k), \tag{6}$$

где

$$g_k = (-1)^k h_{2M-k-1}. \tag{7}$$

Таким образом, многомасштабный вейвлет-анализ сводится к нахождению коэффициентов аппроксимации $s_{j,k}$ и детализирующих коэффициентов $d_{j,k}$ в разложении сигнала $f(t)$ по формуле (2) ортогональным вейвлет-преобразованием.

Выбор материнского вейвлета. При выполнении анализа различных участков нестационарного сигнала с максимальной точностью одним из важнейших решений является выбор вида материнского вейвлета. Общим правилом здесь является то, что вейвлет должен быть похож на форму анализируемого сигнала.

В качестве материнских вейвлетов при получении частотно-временного представления речевого сигнала (сонограмм) выбраны следующие функции: вейвлет Морле; вейвлет Шеннона; вейвлет “мексиканская шляпа”.

К преимуществам данных вейвлетов относится то, что все они задаются точными функциями от времени. При использовании вейвлетов Морле и Шеннона можно, задавая коэффициент масштабирования, изменять их локализацию и достигать высокого частотно-временного разрешения сонограмм. Применение вейвлета “мексиканская шляпа” целесообразно при анализе коротких участков сигнала, так как обеспечивается возможность “рассмотреть” каждый период сигнала в отдельности.

В ходе проведенных исследований при построении изображений “видимый звук” или сонограмм использовалось разработанное специальное программное обеспечение (СПО) WaveView-4, которое является одной из версий ПО семейства вейвлет-анализа сигналов WaveView

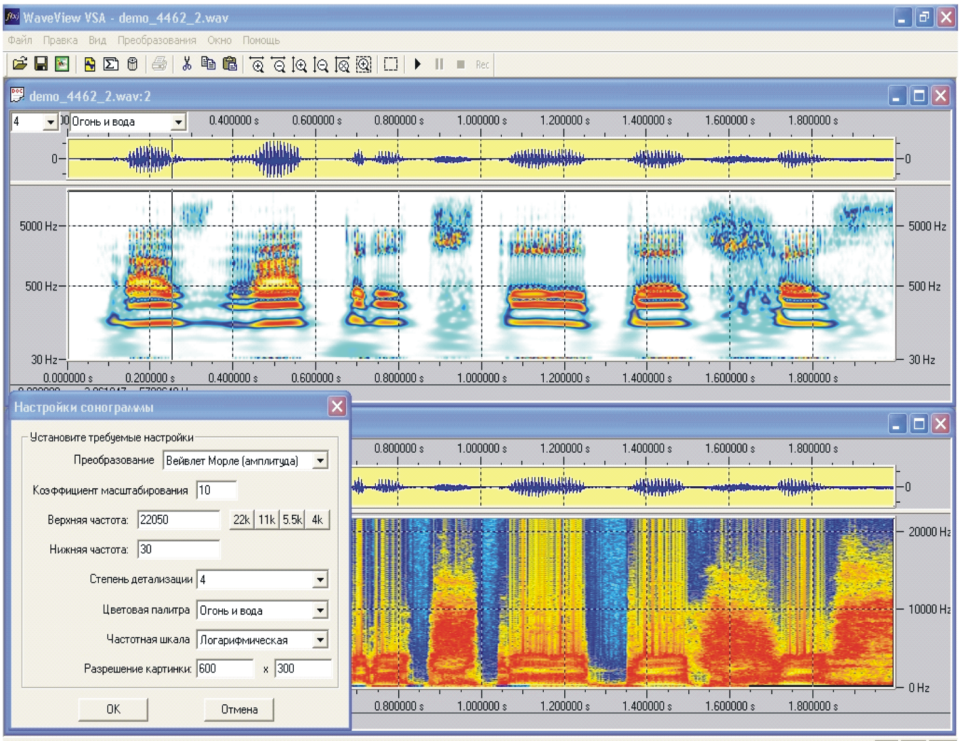

Рис. 3. Сравнение сонограмм. Верхнее изображение — вейвлет-сонограмма; нижнее — фурье-сонограмма; слева внизу — меню СПО WaveView-4

[9]. При получении сонограмм в СПО WaveView-4 применяется алгоритм вычисления непрерывного вейвлет-преобразования [10].

На рис. 3 для сравнения представлены вейвлет- и Фурье-сонограммы слов: “раз”, “два”, “три”, “четыре”, “пять”, “шесть”.

При сравнении сонограмм, полученных на гласных и согласных звуках, заметно существенное преимущество вейвлет-преобразования перед преобразованием Фурье, которое заключается в хорошей локализации как низкочастотных, так и высокочастотных составляющих речевого сигнала.

Исследовательский комплекс “Фон”. На рис. 4 представлена структура исследовательского аппаратно-программного комплекса “Фон”, предназначенного для обработки аудиозаписей и сигналов телефонных переговоров с использованием вейвлет-технологии. Аппаратно-программный комплекс “Фон” построен на основе персонального компьютера с установленным на нем СПО WaveView-4 и подключенными дополнительными устройствами.

Анализ аудиозаписей голосов 351 диктора с использованием СПО WaveView-4 комплекса “Фон” позволяет сделать следующие выводы:

— при построении вейвлет-сонограмм на тональных звуках (гласных) проявляется более сложная, отличающаяся от общеизвестной (формантной), структура речевого сигнала (в области формант F3–F4);

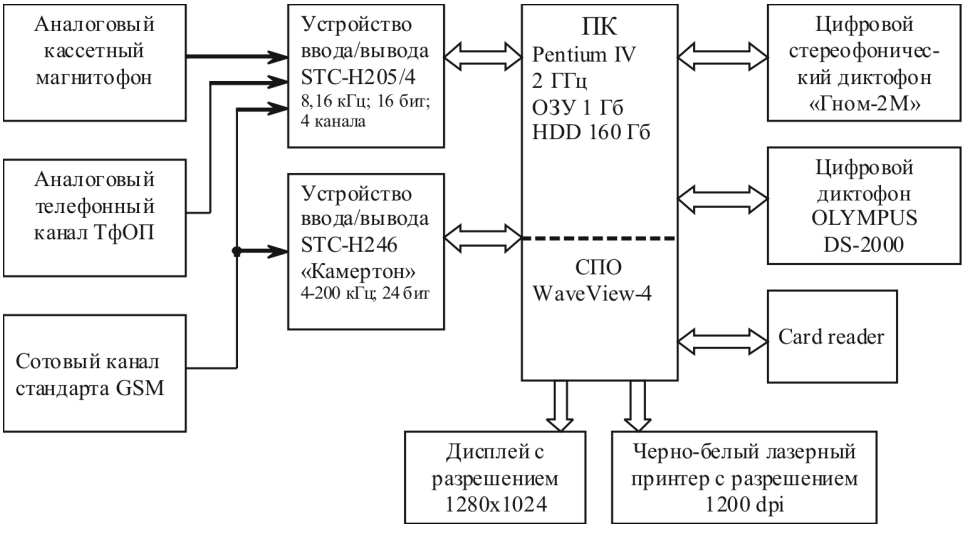

Рис. 4. Структура исследовательского аппаратно-программного комплекса “Фон”

— обеспечивается возможность выделения частотно-временных параметров согласных звуков.

Последний вывод имеет важное практическое значение, так как существующие системы идентификации из всего речевого потока выделяют только тональные участки — гласные: [а], [о], [э], [и], [ы], [у]. Согласные же звуки, щелевые, аффрикаты и сонанты, не используются из-за ограничений, связанных с фурье-преобразованием.

На рис. 5 и 6 представлены для сравнения вейвлет- и фурье-сонограммы слова “шесть”.

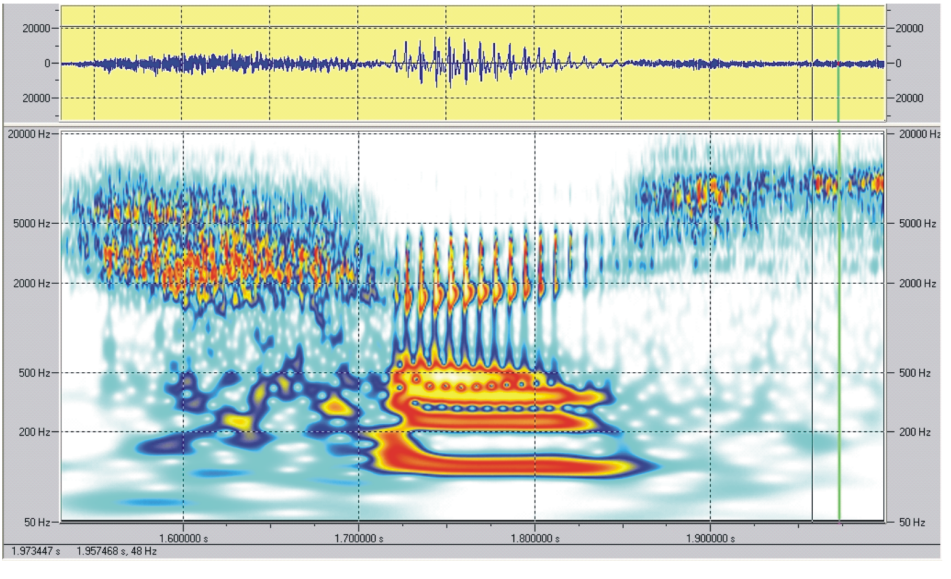

Рис. 5. Вейвлет-сонограмма слова “шесть”

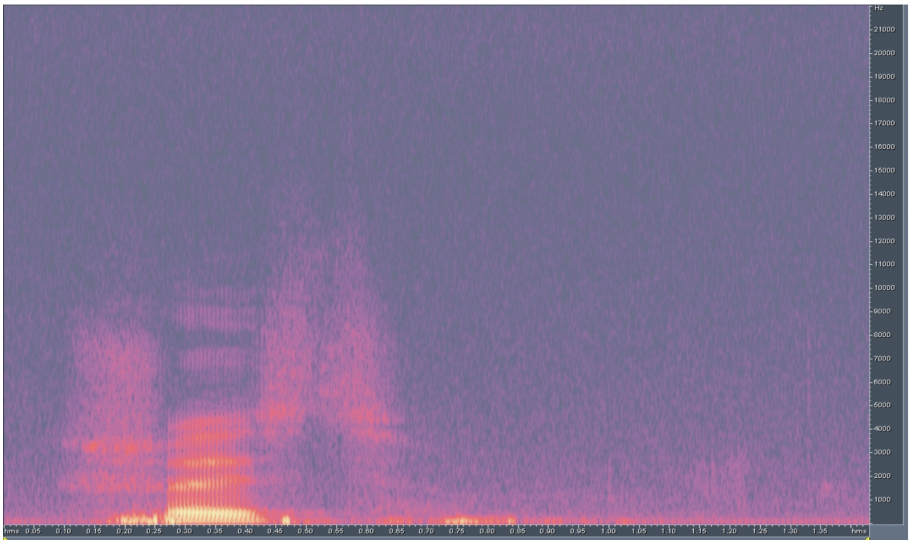

Рис. 6. Фурье-сонограмма слова “шесть”

Заключение. В работе обобщены сведения о составе звуков (гласные, согласные) русского языка, а также наиболее распространенных иностранных языков.

Определены недостатки систем исследования, распознавания и идентификации личности по голосу, использующих алгоритмы анализа гласных звуков, хотя в большинстве языков основная информация передается согласными. Показано, что невозможность получения частотно-временных параметров согласных звуков заложена в использовании преобразования Фурье при обработке нестационарных сигналов.

Разработан исследовательский комплекс частотно-временного анализа речевого сигнала, обеспечивающий получение параметров как гласных, так и согласных звуков. Представлены преимущества вейвлет-технологии по сравнению с фурье-анализом на примерах сравнительного анализа сонограмм слитной речи.

СПИСОК ЛИТЕРАТУРЫ

1. Г а л я ш и н а Е. И. Судебная фоноскопическая экспертиза. – М.: Триада, 2001.
2. Г о р ш к о в Ю. Г. Аппаратно-программные средства анализа, шумоочистки и засекречивания речевого сигнала коммерческого применения (3-е поколение: вейвлет-технологии) // Тез. докл. 1-й Москов. междунар. конф. “Интегрированные системы безопасности: Новейшие технологии”, 26–27 апреля 2004 г. – Москва. С. 2.
3. Ф а н т Г. Акустическая теория речеобразования. (G. Fant. Acoustic theory of speech production, 1960): Пер. с англ. Л.А. Варшавского и В.И. Медведева; Под ред. В.С. Григорьева. – М.: Наука, 1964. – 284 с.

4. Чернов В. Н. Новые программные возможности измерения вероятностных симметричных характеристик акустических сигналов в среде LABVIEW // Образовательные, научные и инженерные приложения в среде LabVIEW и технологии National Instruments. Восьмая Междунар. науч.-практич. конф. – Москва, 2009. – С. 33–35.
5. Тимофеев Е. Н., Голощапова Т. И., Докучаев И. В. Применение автоматизированной системы “Диалект” на базе компьютерной речевой лаборатории CSL (США) при решении задач идентификации дикторов: Учеб. пособие. – М.: ЭКЦ МВД России, 2000. – 120 с.
6. Дьяконов В. И. Вейвлеты: от теории к практике. – М.: Солон-Р, 2002.
7. Новиков Л. В. Основы вейвлет-анализа сигналов: Учеб. пособие. – СПб, ИАнП РАН, 1999. – 152 с.
8. Дремин И. М., Иванов О. В., Нечитайло В. А. Вейвлеты и их использование // Успехи физических наук. – 2001. – Т. 171, № 5. – 2001. – С. 465–500.
9. Горшков Ю. Г., Кузин А. Ю. Применение Wavelet-преобразования при решении задач анализа речевого сигнала // Сб. трудов X Всерос. науч. конф. “Проблемы информационной безопасности в системе высшей школы”. – М., 2003. – С. 24.
10. Горшков Ю. Г., Пестряков А. А. Специализированные средства анализа речевых сигналов с использованием вейвлет-преобразования // Проблемы информационной безопасности в системе высшей школы. XV Всерос. науч.-практич. конф. – Москва, 2008. – С. 36.

Статья поступила в редакцию 22.03.11

Юрий Георгиевич Горшков родился в 1945 г., окончил в 1969 г. Новосибирский электротехнический институт связи. Канд. техн. наук, доцент кафедры “Информационная безопасность” МГТУ им. Н.Э. Баумана. Автор более 40 научных работ в области информационной безопасности и разработки защищенных систем связи.

Y. G. Gorshkov (b. 1945) graduated from the Novosibirsk Electrical Technical Institute of Communication in 1969. Ph. D. (Eng.), assoc. professor of “Information Security” department of the Bauman Moscow State Technical University. Author of more than 40 publications in the field of information security and development of secured telecommunications systems.